

Model for Enhancing Cloud Computing Resource Allocation Management Using Data Analytics

Sello Prince Sekwatlakwatla¹, Vusumuzi Malele²

^{1,2}North-West University, Mafikeng Campus, South Africa
Email: sek.prince@gmail.com¹, vusi.malele@nwu.ac.za²

Abstract

The cloud computing environment requires an adequate and accurate traffic prediction tool to fulfill the needs of customers and support organizations effectively. In the absence of an effective tool for forecasting cloud computing traffic, many organizations might fail. It is difficult to predict the network resources that are suitable to meet the needs of all network clients at a given time in a cloud computing environment because of the inconsistent network traffic flow. There is still room for improving the predictive accuracy of the model in cloud computing. The higher the accuracy of the traffic flow, the better the allocation of resources. Therefore, this study proposes an ensemble method called SGLA (Stepwise Gaussian Linear Autoregressive) by combining linear regression, support vector machines, Gaussian process regression, and the autoregressive integrated moving average technique. SGLA performed better than all methods with a minimum MAPE of 1.03% of the ensemble approach by using the averaging strategy, SGLA shows a clear advantage in handling resource allocation better despite traffic fluctuations, with 95.7% traffic prediction accuracy. Overall experimental results indicate that this method performed better than single models in terms of prediction accuracy. The main contribution of this study is to propose a data analytics model for enhancing cloud computing resource management.

Keywords: ensemble method, resource allocation, traffic prediction

1. INTRODUCTION

Cloud computing has gained popularity due to its ability to simplify and improve the management of ICT resources [1-3]. It can also be used to provide easy, on-demand network access to a large pool of configurable processing resources [4]. In the absence of an effective tool for forecasting cloud computing traffic, many organizations might fail. It is difficult to predict the network resources that are suitable to meet the needs of all network clients at a given time in cloud computing environment because of the inconsistent network traffic flow. For example, it was common for clients to complain about slow system times, application timeouts, and high bandwidth usage because of inconsistent traffic flow [4-5].

A recent study found that some cloud computing providers couldn't allocate enough resources to meet their clients' demands, causing system bottlenecks

during peak periods [5]. Unlike using the single algorithms, combining algorithms and models could have a great deal in predicting cloud computing traffic flow [6]. The combination of models is provided through ensemble method, in ensemble methods, multiple models are combined instead of a single model to improve the accuracy of the results. The ensemble method is aimed at improving the accuracy of results in models by combining multiple models rather than using a single model [6-7], there is relatively little research showing the combination of two or more models or methods with varying applications especially in traffic flow forecasting [8].

Study by [6] propose an ensemble model for short-term traffic prediction; they claim that their model had the highest accuracy among the considered models, but there is no indication of traffic fluctuations. In this paper, a robust ensemble method for predicting short-term traffic flows has been proposed and has improved accuracy. However, the researchers need to keep in mind that abnormal traffic data (such as traffic jams and traffic accidents, which can reflect the basic state of the road's traffic flow) must be considered when formulating an effective forecasting model for short-term traffic flows [7]. By integrating data schemes and deep learning algorithms, this study proposes an ensemble framework for traffic flow prediction. Hybrid traffic flow prediction schemes demonstrated better performance than other models since the RMSE, MAE, and MAPE indicators were at least 30% lower than those of their counterparts, suggesting that hybrid traffic flow prediction schemes performed better. However, the model was tested without considering traffic parameter effects, and thus the accuracy of traffic flow prediction was further validated regardless of traffic parameter influences [8]. By using a deep ensemble neural network model, urban road traffic speed can be accurately predicted, which improves accuracy in traffic state forecasting. The method's limitations include that input variables are based only on traffic flow data and ignore the influence of other factors [9].

An ensemble learning method for long-term prediction of multiple types of time-varying network traffic is presented. The method evaluated prediction accuracy by using both ML metrics (RMSPE) and novel networking metrics. Based on their specific needs, network operators may set parameters regarding overestimation and underestimation, as well as acceptable blocking thresholds. Future work will explore different parameter configurations of the AOBT metric and their effects on the choice of prediction models for multiple types of traffic [10]. An ensemble model based on artificial intelligence for predicting vehicular traffic noise proved to be more robust than single models in dealing with uncertainties when comparing the results obtained by the single models. It is important to note that the linear ensemble techniques in the study have limitations in that a lower performance of one of the single models may result in a less robust single mode performance [11-12]. It improves the generalization ability of cloud computing using at least one of the following strategies: (i) averaging strategy, (ii) voting

strategy, and (iii) learning strategy [10]. As part of this paper, an averaging strategy for predicting traffic is proposed, which is necessary for allocating resources. The method can be easily applied by following these three steps: Develop multiple predictive models that are capable of making their own predictions. Using the same set of training data, train each prediction model, and then average all of the results. Overfitting is less likely to occur in ensemble models because of the diversity in the base model [11]

The Root mean square deviation (RMSD) and the mean absolute percentage error (MAPE) are proposed as methods for evaluating predictive accuracy in this paper. It enables the comparison of the model's accuracy between its predictive value and ensemble value, allowing for direct comparisons between traffic types [12-13]. Against this background, this paper combines linear regression, Gaussian process regression, support vector machine, and autoregressive integrated moving average as a prediction tool. This paper contributes to the ongoing doctoral study that aim to achieve the following objective to propose the data analytics model for enhancing cloud computing resources management.

The methods explain clearly how the author carried out the research. TAs a result of the integration of the Cloud of Things (CoT), a novel linear regression (NLR) methodology can be used to predict energy consumption (EC) and help reduce it significantly, thus creating a smart and sustainable environment. In this model, the assumption is that the model goes along a straight line with a mean point, which is a drawback [1]. Using a support vector machine to analyze the supercritical water heat transfer process, the prediction results are affected by the normalization of the data, and the accuracy rates are 0.5894 before and 0.8247 after normalization of the data [14]. Zhong proposes supporting vector machines for predicting traffic casualties. The method has proven to be an effective tool for overcoming forecast limitations, such as long-running time and difficulty classifying complicated attributes. In the case of one factor considered separately, the effect is poor; when all factors are considered separately, the effect is poor [15-16].

Torotani proposed that the Gaussian Process Regression algorithm be used to predict air traffic controllers' arrivals. The support algorithm can calculate the appropriate indicator near the runway threshold according to the results. However, the indicator has errors in the initial phase of arrival spacing. The speed profile derived from GPR can be applied to the support algorithm to correct these errors [17]. The combination of the autoregressive integral moving average (ARIMA) model and the long short-term memory (LSTM) neural network was proposed, and the results show that the dynamic weighted combined model proposed has a better prediction effect. Moreover, the combined method can combine the advantages of different models to obtain better performance than a single model [18-19]. The rest of this paper is divided into four (4) sections: section 2 focuses

on the method, section 3 deals with results and discussion, and conclusion is briefly presented in Section 4.

2. METHODS

2.1 Data sources

This study uses a hybrid data collection design that includes public (social media), internal (history data), and service data (real-time) from Company X cloud computing resource management on an hourly basis from January 2021 through November 2023, enabling different regression models to be tested. The traffic prediction results using the dataset were also demonstrated using MATLAB simulations.

2.2 Research design

In Figure 1, indicate four different models (linear regression, support vector machines, Gaussian process regression, and autoregressive integrated moving average). Step 1: This layer downloads public, internal, and service data on an hourly basis, and traffic flow prediction is a key part of auto resource allocation. Step 2: involves obtaining and processing data, and a third step involves discovering, cleaning, and synchronizing data issues in Live Editor. Step 3: For traffic prediction, all models will use the same data to forecast traffic on an hourly basis (linear regression, gaussian process regression, support vector machine, and autoregressive integrated moving average). Step 4: The prediction data validation process validated the data and finalized each method's prediction results. Step 5: This layer evaluates the accuracy of all methods.

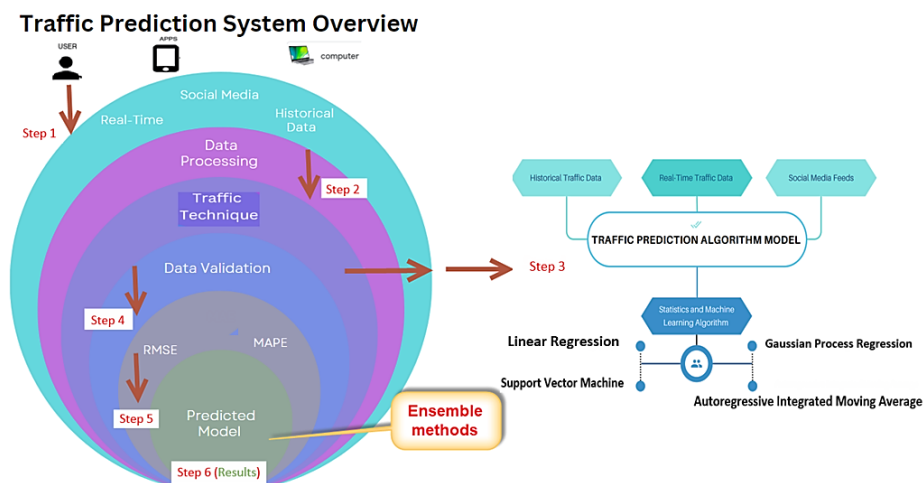


Figure 1. Traffic flow predictive Research Design.

2.2.1 Linear regression

The stepwise regression technique has three basic variants: forward selection, backward elimination, and stepwise regression. Stepwise regression is a special case of hierarchical regression in which statistical algorithms determine the predictors to include in the model [9]. When forward selection is used, a model begins without significant predictors and gradually adds significant predictors until a statistical stopping criterion has been reached. By starting with all possible predictors, the model eliminates non-significant predictors until a stopping criterion is reached [9]. Using the stepwise regression method, predictors are added and removed as the model is being built as shown in Equations 1.

$$P_{j, std} = b_j \left(\frac{\sigma_{x_j}}{\sigma_y} \right) \quad (1)$$

Where σ_y and σ_{x_j} is the normal deviations for the reliant variable and the consistent P_j independent variable.

2.2.2 Support vector machines

SVMs use a subset of training points in the decision function (called support vectors), which makes the method memory efficient [11]. SVMs are used for classification, regression, and outliers' detection. Calculating Support Vector machine predictions are explained by Equations 2 and 3.

$$W(c_i) = \text{sign} \left(\sum_{j=1}^s \alpha_j p_j k(c_j, x_i) + b \right) \quad (2)$$

$$W(v, v') = \exp \left(-\frac{\|v - v'\|^2}{2\gamma^2} \right) \quad (3)$$

Where c_i is the (vector of values) to predict. The c_j se are called support vectors which are a subset of the training data. The p_j re is the class (-1 or +1) of each data p_j . The c_j are constants, one for each p_j . The b is a single numeric constant. Letter s is the number of support vectors. The W is a kernel function that returns a number, 1.0 meaning identical and 0.0 meaning as different as possible, based on the similarity between two vectors.

2.2.3 Gaussian Process Regression

Gaussian processes are composed of finite numbers of random variables with a Gaussian distribution. Which is observation, Gaussian field, and input. The following is the standard formula for the model as shown in Equation 4.

$$J(x) = \sum_{i=1}^n \beta_i f_i(x) + F(x) \quad (4)$$

Where J_i is the reaction, we are interested in, $F(x)$ is a Gaussian process, f_i is known functions, and β is unknown. Suppose there are n sample points x_1, x_2, \dots, x_m , with corresponding sample results y_1, y_2, \dots, y_m , β can then be estimated using the equation. Therefore, we can use it.

$$J(X) = \sum_{i=1}^n \beta_i f_i(X) \quad (5)$$

Then forecast response at sites x .

2.2.4 Autoregressive integrated moving average

ARIMA was first presented by Box & Jenkins in 1970 [1]. Users generally favor the ARIMA model for its accuracy in predicting outcomes and flexibility with various time-series data types. This model combines the AR and MA models, as well as differencing. Regression models (A.R.) are based on past data, whereas moving average models (MA) are based on residuals. The following steps can be used:

$$C_t = \theta_0 + \phi_1 C_{t-1} + \phi_2 C_{t-2} + \dots + \phi_a C_{t-a} + P_t - \phi_1 P_{t-1} - \theta_2 P_{t-2} - \dots - \theta_c E_{t-c} \quad (6)$$

Here, C_t is the actual viewed value at time t and P_t is random error.

$$\phi_1 (I = 1, 2, \dots, a) \text{ and } \theta_a (a=0, 1, 2, \dots, c) \quad (7)$$

There is no meaning or variance in the random errors, and they are both uncorrelated and independent of each other. Model parameters a and c indicate model order. ARIMA (a, b, c) can be expressed in this way, with 'a' indicating the order of the A.R. model and 'b' indicating the degree of differentiation. In the process of ARIMA model recognition, parameter selection, and verification, the parameters listed above are iteratively determined. Before identifying the time series, it is important to check its stationarity. ARIMA models are designed for stationary time series. If it is not stationary, differencing can be used to make the series stationary. Second, the A.R. and M.A. models are fitted according to autocorrelation function and partial autocorrelation function plots based on the stationary data.

2.2.5 Ensemble Method for Traffic prediction

Linear regression, support vector machines, Gaussian process regression, and the autoregressive integrated moving average technique model are combined to form SGLA (Stepwise Gaussian Linear Autoregressive) as a proposed method see figure 2. The goal is to improve the accuracy of results in models by combining multiple models instead of using just one.

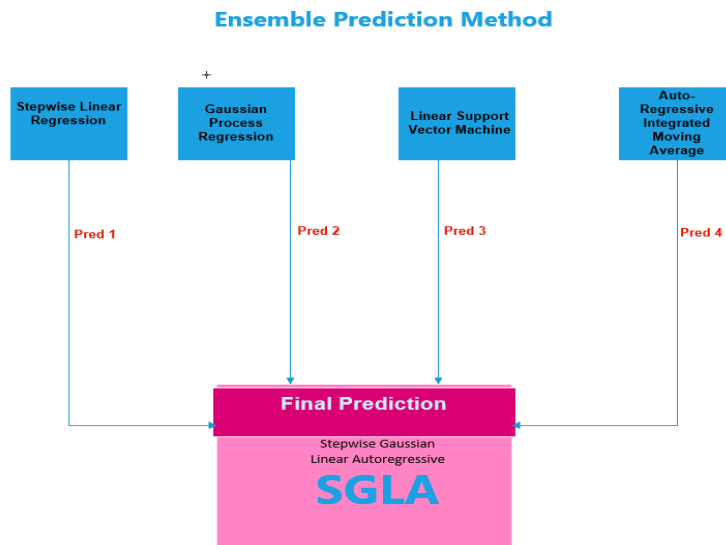


Figure 2. Ensemble Method

2.3 Evaluation Measures

The Root mean square deviation (RMSD) and the mean absolute percentage error (MAPE) are proposed as methods for evaluating predictive accuracy. It enables the comparison of the model's accuracy between its predictive value and ensemble value, allowing for direct comparisons between traffic types [16-17]. RMSD and MAPE as shown in Equation 8 and 9.

$$RMSD = \sqrt{\frac{\sum_{i=1}^R (C_i - \hat{C}_i)^2}{R}} \quad (8)$$

Where i = variable 1, R Number of data points with no missing values, C_i is the real observations time series and \hat{C}_i the time series is estimated.

$$MAPE = \frac{100}{n} \sum_{t=1}^n \frac{F_t - P_t}{F_t} \quad (9)$$

Where F_t is the actual value and P_t is the predicted value.

3. RESULTS AND DISCUSSION

3.1. Experimental Results

In this meticulously designed experiment, the division of collected data was strategically allocated with 90% dedicated to training and the remaining 10% reserved for testing. To ensure the highest data integrity for the simulations carried out in Matlab, a rigorous preprocessing stage was implemented. This stage involved the removal of outliers and the imputation of missing values, setting a solid foundation for the analysis that followed. The experiment's outcomes, depicted across various figures, offer insightful revelations into the performance and characteristics of different predictive models. Figure 3 illustrates the performance of the Linear Regression model, noting an upload time spanning from 19 to 38 seconds and a modest model size of 27 kb. Remarkably, this model achieved a 91% prediction accuracy while operating under conditions of progressively increasing traffic flow, demonstrating both efficiency and reliability.

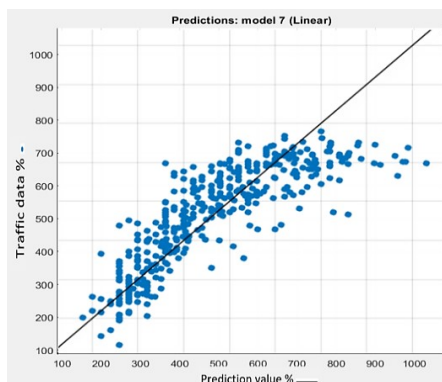


Figure 3. Linear regression

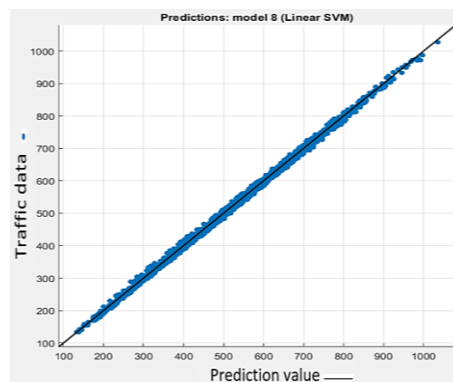


Figure 4. Support vector machines

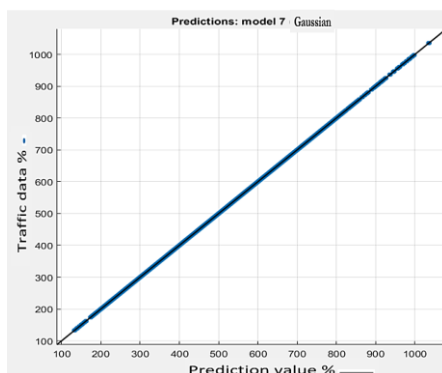


Figure 5. Gaussian Process Regression

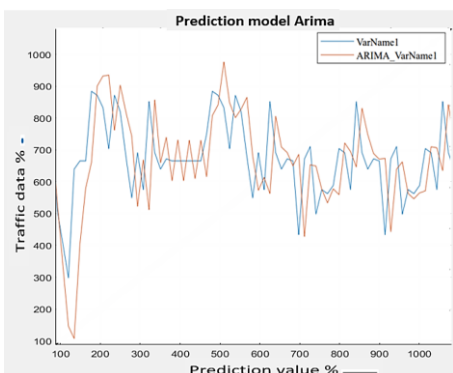


Figure 6. Autoregressive integrated moving average

In contrast, Figure 4 delves into the Support Vector Machine (SVM) model, which exhibited upload times ranging from 30 to 38 seconds and a model size of 40 kb. The processing time for the SVM model was notably longer than that of the Linear Regression model, despite facing the same traffic conditions, as indicated in subsequent figures. Yet, the SVM model upheld a commendable accuracy rate of 90.2%, with an observed increase in traffic flow during its simulation, underscoring its robustness amidst varying conditions. Turning attention to Figure 5, the Gaussian Process Regression model is highlighted, demanding a significantly longer upload time of 490 seconds and a larger model size of 580 kb. This model, despite its extensive processing time compared to its predecessors, outshone them with a superior accuracy rate of 92.5%. This finding underscores the model's enhanced predictive capabilities, even in the face of more substantial data and computational demands.

Lastly, Figure 6 presents the Autoregressive Integrated Moving Average model, which recorded an upload time of 380 seconds and the largest model size of 650 kb among the models evaluated. Interestingly, while its processing time was quicker in comparison, the model's prediction accuracy was observed at 78.2%, the lowest among the models tested. This outcome hints at the intricate balance between model complexity, processing efficiency, and predictive accuracy, inviting further exploration into the optimization of such models for real-world applications. Together, these findings weave a comprehensive narrative on the efficacy, efficiency, and adaptability of different predictive modeling techniques in handling dynamic data streams, setting a benchmark for future explorations in the field.

3.2 Ensemble Methods

The ensemble method is aimed at improving the accuracy of results in models by combining multiple models rather than using a single model [14-15], there is relatively little research showing the combination of two or more models or methods with varying applications especially in traffic flow forecasting [4-6]. It improves the generalization ability of cloud computing using at least one of the following strategies: As part of this paper, an averaging strategy for predicting traffic is proposed, which is necessary for allocating resources. The method can be easily applied by following these three steps: Develop multiple predictive models that are capable of making their own predictions. Using the same set of training data, train each prediction model, and then average all of the results. Overfitting is less likely to occur in ensemble models because of the diversity in the base model [15].

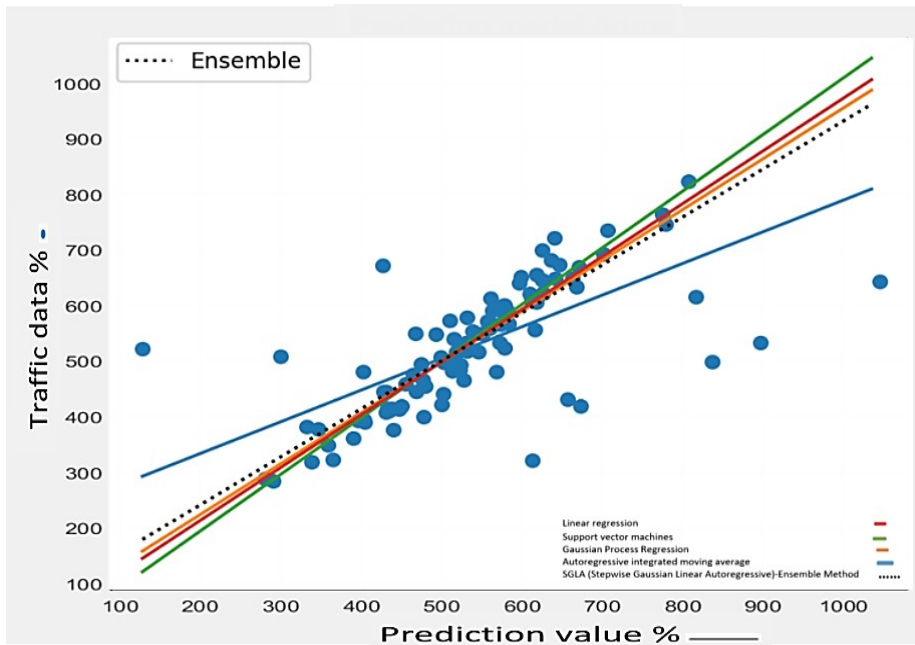


Figure 7. Ensemble results

Figure 7 gives a visualization of the distances between the results obtained by considered methods across traffic predictions, according to each prediction method. As seen in the figure, the difference between individual prediction results and ensemble predictions is statistically significant, the ensemble results are shown in Table 1, and the results of SGLA (Stepwise Gaussian Linear Autoregressive) are shown in broken lines, which demonstrate the model's accuracy better than any other model despite traffic fluctuations.

Table 1. Overall forecasting results

Methods	Rmsd (%)	MaPe (%)	Prediction (%)
Linear regression	1.4	1.6	94.1
Support vector machines	2.4	2.1	91.2
Gaussian Process Regression	1.7	1.9	92.5
Autoregressive integrated moving average	2.5	2.2	89.9
SGLA (Stepwise Gaussian Linear Autoregressive)-Ensemble Method	1.3	1.05	95.7

To aggregate the obtained results and check if differences between the methods are statistically significant. In terms of RMSD, Linear regression and Gaussian Process Regression perform better with minimum errors of 1.4 % and 1.7 % respectively, Support vector machines is having error of 2.4 % and Autoregressive integrated moving average 2.5% however SGLA performed better than all method with minimum of MAPE with 1.05 % of the ensemble approach by using averaging strategy, followed by Linear regression with 1.6 % , Gaussian Process Regression MAPE errors is 1.9 %, Support vector machines and Autoregressive integrated moving average error where 2.1 % and 2.2 respectively, Linear regression predictions is 94.1 % however model is having high errors processing the traffic in the initial phase, Support vector machines and Gaussian Process Regression perform better with 91.2 % and 92.5 % and Autoregressive integrated moving average is 89.9% however when the traffic increase the model decrease the accuracy results, model is unable to handle the high volume of traffic, SGLA showing a clear advantage of handling the resources allocation better despite traffic fluctuations with 95.7% traffic prediction accuracy.

4. CONCLUSION

This study proposes an ensemble method called SGLA (Stepwise Gaussian Linear Autoregressive) by combining linear regression, support vector machines, Gaussian process regression, and the autoregressive integrated moving average technique. SGLA performed better than all methods with a minimum MAPE of 1.03% of the ensemble approach by using the averaging strategy, followed by linear regression with 1.6%, Gaussian process regression MAPE errors of 1.9%, support vector machines and autoregressive integrated moving average error of 2.1% and 2.2%, respectively. Linear regression outperforms all models with 94.1%, but the model is having high errors processing the traffic in the initial phase. Support vector machines and Gaussian Process Regression perform better with 91.2% and 92.5%, respectively. However, when traffic increases, the model's accuracy decreases, making it unable to handle the high volume of traffic. SGLA shows a clear advantage in handling resource allocation better despite traffic fluctuations, with 91.7% traffic prediction accuracy. Experimental results indicate that this method performed better than single models in terms of prediction accuracy. This paper contributes the analytics model for enhancing cloud computing traffic flow and resource management.

REFERENCES

- [1]. N. K. Biswas, S. Banerjee, U. Biswas, and U. Ghosh, "An approach towards development of new linear regression prediction model for reduced energy consumption and SLA violation in the domain of green cloud computing,"

- Journal of Sustainable Energy Technologies and Assessments*, vol. 45, 2021, doi: 10.1016/j.seta.2021.101087
- [2]. M. Dongliang, L. Yi, Z. Tao, H. Yanping. "Research on prediction and analysis of supercritical water heat transfer coefficient based on support vector machine," *Journal of Nuclear Engineering and Technology*, Vol. 55, no. 11, pp. 4102-4111, 2023, doi: 10.1016/j.net.2023.07.030
- [3]. D. Toratani, T. Yoshihara and A. Senoguchi, "Support algorithm for air traffic controllers' arrival spacing: Improvement of trajectory estimation using Gaussian Process Regression," *Journal of Control Engineering Practice*, vol.128, 2022, doi: 10.1016/j.conengprac.2022.105343.
- [4]. W. Zhong and L. Du, "Predicting Traffic Casualties Using Support Vector Machines with Heuristic Algorithms: A Study Based on Collision Data of Urban Roads," *Journal of Machine Learning and Big Data Analytics for Sustainability and Resilience*, vol, 15, 2023, doi: 10.3390/su15042944
- [5]. S. Lu, Q. Zhang, G. Chen, and D. Seng: "A combined method for short-term traffic flow prediction based on recurrent neural network", *Alexandria Engineering Journal*, vol. 60, pp. 87–94, 2021. doi: 10.1016/j.aej.2020.06.008
- [6]. G. Zheng, W. K. Chai, V. Katos and M. Walton. "A joint temporal-spatial ensemble model for short-term traffic prediction," *Journal of Neurocomputing*, vol. 457, pp.26–39, 2021. doi: 10.1016/j.neucom.2021.06.028
- [7]. H. Yan, L. Fu, Q. Yong, and Y. Dong-Jun, "Robust ensemble method for short-term traffic flow prediction," *Journal of Future Generation Computer Systems*, vol. 133, pp. 395–410, 2022. doi: 10.1016/j.future.2022.03.034
- [8]. X. Chen et al., "Traffic flow prediction by an ensemble framework with data denoising and deep learning model," *journal of Physica*, vol. 565, 2021. doi: 10.1016/j.physa.2020.125574
- [9]. L. Wenqi et al., "Traffic speed forecasting for urban roads: A deep ensemble neural network model," *Journal of Physica*, vol. 593, 2022. doi: 10.1016/j.physa.2022.126988
- [10]. A. Knapieńska et al., "Long-term prediction of multiple types of time-varying network traffic using chunk-based ensemble learning," *Journal of Applied Soft Computing*, vol.130, 2022. doi: 10.1016/j.asoc.2022.109694
- [11]. V. Nourania, H. Gökçekuşb and I. K. Umarb, "Artificial intelligence-based ensemble model for prediction of vehicular traffic noise," *journal of Environmental Research*, vol.180, 2020. doi: 10.1016/j.envres.2019.108852
- [12]. J. Kamiri, and G. Mariga, "Research Methods in Machine Learning: A Content Analysis," *International Journal of Computer and Information Technology*, vol. 10, no. 2, 2021.
- [13]. Zhou et al., "Genetic Algorithm with Heuristic-based Local Search for multi-dimensional resources scheduling of cloud computing," *journal of Applied Soft Computing*, vol.136, 2023. doi: 10.1016/j.asoc.2023.110027
- [14]. F. Li, W. Ma, H. Li and J. Li, "Improving Intrusion Detection System Using Ensemble Methods and Over-Sampling Technique," *2022 4th International Academic Exchange Conference on Science and Technology Innovation (IAECST)*,

- Guangzhou, China, pp. 1200-1205, 2022, doi: 10.1109/IAECST57965.2022.10062178.
- [15]. T. Ahmad and N. Zhou, "Ensemble Methods for Probabilistic Solar Power Forecasting: A Comparative Study," *2023 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1-5, 2023. doi: 10.1109/PESGM52003.2023.10253133.
- [16]. Feng et al, "An ensemble machine learning approach for classification tasks using feature generation," *Journal of Connection Science*, Vol. 35, 2023 doi: 10.1080/09540091.2023.2231168.
- [17]. A. B. Ismail, H. B. A. Bakar, and S. B. Shafei, "Comparison of LDPE/corn stalk with eco degradant and LDPE/corn stalk with MAPE: Influence of coupling agent and compatibiliser on mechanical properties," *Materials Today: Proceedings*, vol. 31, pp. 360-365, 2020.
- [18]. A. V. Agranovskii and A. P. Silukov, "Comparative Analysis of Results of Modern Classification Algorithms Usage for Determining the Type of Physical Activity Based on Integrated Sensors Data," *Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, pp. 1-5, 2021 doi: 10.1109/WECONF51603.2021.9470661.
- [19]. Zhang et al., "Security computing resource allocation based on deep reinforcement learning in serverless multi-cloud edge computing," *journal of Future Generation Computer Systems*, vol. 151, pp. 152-161, 2024. doi: 10.1016/j.future.2023.09.016.